

# Stochastic Comparisons between first-passage times for Markov Chains.

Emilio De Santis

Department of Mathematics  
La Sapienza University of Rome  
desantis@mat.uniroma1.it

Fabio Spizzichino

Department of Mathematics  
La Sapienza University of Rome  
fabio.spizzichino@uniroma1.it

October 4, 2012

## Abstract

We develop some sufficient conditions for the stochastic ordering between first-passage times, in a fixed state, for two Markov chains. In particular, we focus attention on the so called skip-free Markov chains. For our purposes, we develop a special type of coupling. We also define a relation between two Markov chains, which can have a natural role in the comparison between the tail behaviors of the distributions of first-passage times. Finally, we present some examples dedicated to words' occurrences.

**Key Words:** Skip-free Markov chains, Coupling, Asymptotic Stochastic Comparison, Word Occurrences, Leading Numbers.

2010 Mathematics Subject Classification: Primary: 60E15, 60J10

## 1 Introduction

We consider a Markov chain  $\mathbf{X} \equiv \{X_n\}_{n=0,1,\dots}$  on the state space  $E = E_k \equiv \{0, 1, \dots, k\}$  or  $E = E_\infty \equiv \{0, 1, \dots\}$ . We denote by  $T_h$  the stopping times

$$T_h = \inf\{n \in \mathbb{N} : X_n \geq h\}, \quad h = 1, 2, \dots, k. \quad (1)$$

$T_h$  is thus the random time needed to reach or exceed the *level*  $h$ . In particular we will consider transition matrices  $P = (p_{i,j})_{i,j \in E}$  with the property

$$p_{i,j} = 0 \quad \text{if } 1 \leq i+1 < j. \quad (2)$$

The literature devoted to this topic is very wide. See e.g. [10, 8, 5] and references cited therein. Besides the theoretical interest, the analysis of the first-passage times  $T_h$  for this class of Markov chains emerges in the applications of probability to different fields such as reliability, networks, biology, and so on. See also [1] for some related discussion (Success Runs and Machine

Replacement). In particular chains of this type are encountered, in a fairly direct way, in the problem of first occurrences of words in random sequences of letters from an alphabet. In such a case we have  $E = E_k$ . For simplicity sake, from now on we will limit our attention on the finite cases  $E = E_k$ . However most of our results can be appropriately extended to the infinite state-space case. Throughout the paper, we will denote by  $\Upsilon_k$  the class of transition matrices on the state space  $E_k$  satisfying (2). With some abuse of notation we also say that a Markov chain  $\mathbf{X}$  is in  $\Upsilon_k$  if its transition matrix is in  $\Upsilon_k$ .

The problem of words' occurrence suggested some of our results and will be briefly recalled in the last section. A very large literature has been devoted to such a field, in different frameworks and from different points of view; typically attention has been concentrated on different aspects of the exact computation of  $\mathbb{E}(T_k)$  or of the probability distribution of  $T_k$ .

In this paper we rather consider, for pairs of Markov chains  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$  on the same state space  $E_k$ , stochastic orderings between the corresponding first-passage times  $T_k$  and  $\tilde{T}_k$ . The idea of studying stochastic ordering of first-passage times was already considered in the papers [5] and [10]. In particular, the paper by Irle and Gani, i.e. [10], presents some results in the same spirit of ours, in the context of detection of words.

Different notions of stochastic orders might be considered for the  $\mathbb{N}$ -valued random variables  $T_k$  and  $\tilde{T}_k$  (see e.g. [16, 17]); as natural ones in our context, we consider the *usual stochastic order*  $T_k \preceq_{st} \tilde{T}_k$  and the tail (or asymptotic) stochastic order, that is defined in terms of tail behavior of the distributions. A rather detailed analysis of the stochastic tail order has been offered in the recent work [9].

In our results concerning the stochastic order, the assumption that  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$  belong to  $\Upsilon_k$  will be specifically used. The proof of our results in such direction will be based on a coupling method that takes essentially into account the order structure of the state space  $E_k$ . More precisely, on a same probability space, we construct two Markov chains (sharing the laws of  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$ , respectively) in such a way that they are “coupled” only in some instants when they visit the same states. When one of the two chains has a transition to a state “higher” than the other one, it stops and waits for the latter, which has an independent evolution in the meantime. A similar approach has been also used in [7]. We believe that such a type of (partial) coupling can have some more general interest and that it might be applied in different other problems related with stochastic comparisons. Furthermore we point out that this method of coupling turns out to be a constructive one and can work especially well in the proof of strict inequalities between expected values of interest.

For the results concerning the asymptotic stochastic order, we use different methods of proofs. In such a frame, we need to define a suitable order relation between two stochastic matrices of the same size. The circumstance that such a relation is maintained under products will have a relevant role in our derivations. In this part of the paper the condition (2) will not be necessary.

Our results may also be used to deal with the case of continuous time. Processes in continuous time with a property analogous to (2) have been called *free of positive skips* (see [11]). In [13] it has been shown that, under simple conditions, first-passage times for such processes have the *New Better than Used* property. It is simple to see that also in our case the first-passage times  $T_k$  have the New Better than Used property in discrete time.

In the specific cases of chains related to word occurrences, the distributions of  $T_k$ 's are *generalized geometric distributions* (see e.g. [6]). From a distributional point of view, such discrete distributions may appear rather simple at a first glance. In particular, similarly to the geometrical ones, they are completely determined by their expected values. However they manifest several apparently paradoxical aspects (see in particular [2, 3, 4]). On the basis of our analysis one might show that some unexpected behavior emerge also in the comparison between inequalities of the type  $T_k \preceq_{st} \tilde{T}_k$  and those of the type  $P(T_k \leq \tilde{T}_k) \geq P(\tilde{T}_k \leq T_k)$ .

Several results in the literature concerning waiting times to words' occurrences have been based on the notion of *leading number* associated to a word. Such an analysis is, in a sense, alternative to the one based on Markov chains. As a main feature of this paper, we discuss that the two different approaches can be usefully compared and combined.

The structure of the paper is as follows. In Section 2 we present our results concerning the stochastic order. Section 3 will be devoted to the case of the asymptotic stochastic order. In Section 4 we discuss the theme of waiting times to words' occurrences and present, on such a basis, some examples of applications and some remarks concerning with the results of the previous Sections 2 and 3.

## 2 A class of Markov chains and stochastic comparisons between absorbing times

First we recall the standard notation for the usual stochastic ordering between two real random variables. For  $X, Y$ ,  $X : \Omega \rightarrow \mathbb{R}$ ,  $Y : \Omega' \rightarrow \mathbb{R}$ , we write  $X \succeq_{st} Y$  if the following two equivalent conditions hold

- a)  $P(X > t) \geq P(Y > t)$  for any  $t \in \mathbb{R}$ ;
- b)  $\mathbb{E}[g(X)] \geq \mathbb{E}[g(Y)]$  for all increasing functions  $g : \mathbb{R} \rightarrow \mathbb{R}$  for which the expectations  $\mathbb{E}[g(X)]$ ,  $\mathbb{E}[g(Y)]$  exist.

We will use the same symbol  $\preceq_{st}$  also to compare two probability distributions over  $\mathbb{R}$ .

Notice that the relation  $X \succeq_{st} Y$  does not require that  $X, Y$  are defined on a same probability space; however we recall the following important characterization (see e.g. [16]), that will be used in what follows:  $X \succeq_{st} Y$  if and only if there exists a probability space  $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$  and two random variables  $\hat{X} : \hat{\Omega} \rightarrow \mathbb{R}$ ,  $\hat{Y} : \hat{\Omega} \rightarrow \mathbb{R}$  such that

- $X =^d \hat{X}$ ,  $Y =^d \hat{Y}$ ;
- $P(\hat{X} \geq \hat{Y}) = 1$ .

In what follows we present several results in which we establish stochastic comparisons between first passage times of two different Markov chains in  $\Upsilon_k$ .

We consider two Markov chains  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$ , belonging to  $\Upsilon_k$ , with transition matrices  $P = (p_{i,j})_{i,j \in E}$ ,  $\tilde{P} = (\tilde{p}_{i,j})_{i,j \in E}$ , and initial distributions  $\pi_0 = (\pi_0(i))_{i \in E}$ ,  $\tilde{\pi}_0 = (\tilde{\pi}_0(i))_{i \in E}$ , respectively. We furthermore consider  $T_h$  and  $\tilde{T}_h$  where  $T_h$  is defined for  $\mathbf{X}$  in (2) and  $\tilde{T}_h$  is the analogue for  $\tilde{\mathbf{X}}$ .

In our first result we exploit the condition that the Markov chains belong to  $\Upsilon_k$ . The proof is based on the coupling method, namely on the characterization of  $\preceq_{st}$  given above. However, we do not require the existence of a coupling for two chains having their trajectories almost surely ordered.

**Theorem 1.** *Let  $P = (p_{i,j} : i, j = 0, \dots, k)$  and  $\tilde{P} = (\tilde{p}_{i,j} : i, j = 0, \dots, k)$  be two transition matrices in  $\Upsilon_k$ . Assume that, for any  $i = 0, \dots, k-1$ , there exists  $m(i)$  such that*

- i)  $i + m(i) \leq k$ ;
- ii)  $p_{i,\cdot}^{(m(i))} \succeq_{st} \tilde{p}_{i,\cdot}^{(m(i))}$ .

Moreover suppose that the initial measures are stochastically ordered  $\pi \succeq_{st} \tilde{\pi}$ . Then

$$T_k \preceq_{st} \tilde{T}_k. \quad (3)$$

*Proof.* Let us fix a particular choice of  $m(1), \dots, m(k-1)$  such that i) and ii) hold. Moreover, for future convenience, we fix  $m(k) = 1$ . We will use a coupling method and we will obtain the proof in a recursive way.

On a same probability space  $(\Omega, \mathcal{F}, P)$ , we define a sequence of i.i.d. random variables  $\mathbf{U} = \{U_n\}_{n \in \mathbb{N}}$  and an independent array of i.i.d. random variables  $\tilde{\mathbf{U}} = \{\tilde{U}_{k,n}\}_{k \in \mathbb{N}, n \in \mathbb{N}_+}$ . All these variables have uniform distribution on  $[0, 1]$ .

By using  $\mathbf{U}$ , we construct on  $(\Omega, \mathcal{F}, P)$  a homogeneous Markov chain  $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$  having the law given by the initial distribution  $\pi = (\pi_0, \dots, \pi_k)$  and transition matrix  $P = (p_{i,j})_{i,j \in E_k}$ . We will also construct, by using  $\mathbf{U}$  and  $\tilde{\mathbf{U}}$ , a homogeneous Markov chain  $\tilde{\mathbf{X}} = (\tilde{X}_n)_{n \in \mathbb{N}}$  having the law given by the initial distribution  $\tilde{\pi} = (\tilde{\pi}_0, \dots, \tilde{\pi}_k)$  and transition matrix  $\tilde{P} = (\tilde{p}_{i,j})_{i,j \in E_k}$ . We will prove that the stopping times  $T_k$  and  $\tilde{T}_k$ , corresponding to  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$  respectively, are ordered in the sense that

$$T_k(\omega) \leq \tilde{T}_k(\omega),$$

for each  $\omega \in \Omega$ .

First, we define  $X_0$  and  $\tilde{X}_0$  with distribution  $\pi$  and  $\tilde{\pi}$ , respectively.

We set

$$X_0(U_0) := \inf\{i \leq k : \sum_{l=0}^i \pi_l \geq U_0\}, \quad (4)$$

and analogously

$$\tilde{X}_0(U_0) := \inf\{i \leq k : \sum_{l=0}^i \tilde{\pi}_l \geq U_0\}. \quad (5)$$

It is immediately seen that  $X_0(U_0) \sim^d \pi$  and  $\tilde{X}_0(U_0) \sim^d \tilde{\pi}$ . Furthermore, for each value  $u \in [0, 1]$   $X_0(u) \geq \tilde{X}_0(u)$ , in view of the assumption  $\pi \succeq_{st} \tilde{\pi}$ .

Letting  $I(0) = I(0; U_0) := X_0(U_0)$ , and recalling the meaning of  $m(1), \dots, m(k-1), m(k) = 1$ , we recursively define, for  $n = 1, 2, \dots$

$$X_{m(I(0))+\dots+m(I(n-1))} = X_{m(I(0))+\dots+m(I(n-1))}(U_0, \dots, U_n) := \inf\{s \leq k : \sum_{l=0}^s p_{I(n-1),l}^{(m(I(n-1)))} \geq U_n\}, \quad (6)$$

$$I(n) = I(n; U_0, \dots, U_n) := X_{m(I(0))+\dots+m(I(n-1))}. \quad (7)$$

We notice that, if for a given  $n$ , we obtain  $I(n) = k$  then also

$$I(n+1) = k \text{ and } X_{m(I(0))+\dots+m(I(n-1))+1} = k.$$

We claim that

$$T_k = \sum_{r=1}^L m(I(r)), \quad (8)$$

where  $L$  is the random index

$$L := \inf\{l \in \mathbb{N} : I(l) = k\}. \quad (9)$$

We stipulate that the sum (8) is equal to zero if  $L = 0$ . It is clear that, when  $L = 0$ , (8) holds. Now we prove that also in the case  $L > 0$  the expression for  $T_k$  given in (8) holds true. In fact, at least the inequality  $T_k \leq \sum_{r=1}^L m(I(r))$  holds, since  $I(L) = X_{\sum_{r=1}^L m(I(r))} = k$ .

We then want to show that for any  $t < \sum_{r=1}^L m(I(r))$  one has  $X_t \neq k$ . Let us first consider the values  $a_s = \sum_{r=1}^s m(I(r))$  with  $s = 1, \dots, L-1$ . For these values,  $X_{a_s} = k$  would contradict the position (9).

Let us then consider the discrete intervals of the form  $B_s = \{a_s + 1, \dots, a_{s+1} - 1\}$ , with  $s \in \{1, \dots, L-1\}$  and such that  $a_s + 1 \leq a_{s+1} - 1$ . For  $a \in B_s$ , it is impossible that  $X_a = k$ . In fact  $a + m(a) \leq k$  for  $a \leq k-1$  and  $X_c - X_b \leq (c-b)$  for any  $c \geq b$ .

We now proceed to construct  $\tilde{T}_k$ . To this purpose we consider a sequence of independent Markov chains  $\{\tilde{\mathbf{Y}}^{(r)}\}_{r \in \mathbb{N}}$ . For any  $r = 0, 1, \dots$ , the Markov chain  $\tilde{\mathbf{Y}}^{(r)} = \{\tilde{Y}_n^{(r)}\}_{n \in \mathbb{N}}$  will be such that  $\tilde{Y}_0^{(r)} = 0$  with probability one and it will admit  $\tilde{P}$  as transition matrix. More precisely  $\tilde{\mathbf{Y}}^{(r)}$  is constructed in terms of  $\{\tilde{U}_{r,1}, \tilde{U}_{r,2}, \dots\}$  as follows: for  $n = 1, 2, \dots$

$$I^{(r)}(n-1) := \tilde{Y}_{n-1}^{(r)}(\tilde{U}_{r,1}, \dots, \tilde{U}_{r,n-1}), \quad (10)$$

$$\tilde{Y}_n^{(r)}(\tilde{U}_{r,1}, \dots, \tilde{U}_{r,n}) := \inf\{s \leq k : \sum_{l=0}^s \tilde{p}_{I^{(r)}(n-1),l} \geq \tilde{U}_{r,n}\}. \quad (11)$$

As a function of  $U_0, U_1, \dots$ , we now also define the sequence  $\tilde{\mathbf{Y}} = \{\tilde{Y}_n\}_{n \in \mathbb{N}}$  as follows:

$$\tilde{Y}_0 = \tilde{Y}_0(U_0) := \tilde{X}_0(U_0) \quad (12)$$

$$\tilde{Y}_n(U_0, \dots, U_n) := \inf\{s \leq k : \sum_{l=0}^s \tilde{p}_{I(n-1),l}^{(m(I(n-1)))} \geq U_n\}. \quad (13)$$

Notice that the random variables  $I(0), I(1), \dots$  appearing in r.h.s. of (13) have been defined in (7). Furthermore we have, by construction,  $\tilde{Y}_n(U_0, \dots, U_n) \leq I(n; U_0, \dots, U_n)$  in view of condition ii). The sequences  $\tilde{\mathbf{Y}}$  and  $\{\tilde{\mathbf{Y}}^{(r)}\}_{r \in \mathbb{N}}$  are stochastically independent.

Let now, for  $r \in \mathbb{N}$ ,

$$N_1^{(r)} := \inf\{n \in \mathbb{N} : \tilde{Y}_n^{(r)} = \tilde{Y}_r\}, \quad (14)$$

$$N_2^{(r)} := \inf\{n \in \mathbb{N} : \tilde{Y}_n^{(r)} = I(r)\}. \quad (15)$$

We notice that, for any  $r = 0, 1, 2, \dots$ ,  $N_1^{(r)} \leq N_2^{(r)}$  since the chain  $\tilde{\mathbf{Y}}^{(r)}$  starts in zero, it increases at most of one unit at any step, and  $\tilde{Y}_r \leq I(r)$ . For any  $r \in \mathbb{N}$ ,  $N_1^{(r)}$  and  $N_2^{(r)}$  are two stopping times with respect to the filtration  $(\mathcal{F}_n^{(r)})_{n \in \mathbb{N}}$  where  $\mathcal{F}_0^{(r)} = \sigma(\tilde{Y}_r, I(r))$  and

$$\mathcal{F}_n^{(r)} = \sigma(\tilde{Y}_r, I(r), \tilde{U}_{r,1}, \dots, \tilde{U}_{r,n}) \text{ for any } n = 1, 2, \dots$$

Now we consider the random variables

$$Z_r := \sum_{n=0}^r (N_2^{(n)} - N_1^{(n)}) + \sum_{n=0}^r m(I(n)), \quad (16)$$

for  $r = 1, \dots, L$  and

$$\tilde{X}_{Z_r} := \tilde{Y}_r. \quad (17)$$

Notice that, letting  $r = L$  in (16), one has

$$Z_L := \sum_{n=0}^L (N_2^{(n)} - N_1^{(n)}) + \sum_{n=0}^L m(I(n)) = \sum_{n=0}^L (N_2^{(n)} - N_1^{(n)}) + T_k, \quad (18)$$

Furthermore, by recalling definition (14),  $\tilde{X}_{Z_r} = \tilde{Y}_{N_1^{(r)}}^{(r)}$ .

We now consider the following sequence of random variables:

$$\tilde{Y}_0, \tilde{Y}_{N_1^{(0)}+1}^{(0)}, \dots, \tilde{Y}_{N_2^{(0)}}^{(0)}, \tilde{Y}_1, \tilde{Y}_{N_1^{(1)}+1}^{(1)}, \dots, \tilde{Y}_{N_2^{(1)}}^{(1)}, \tilde{Y}_2, \quad (19)$$

obtained by gluing together the sections of trajectories

$$\tilde{Y}_0; \tilde{Y}_{N_1^{(0)}+1}^{(0)}, \dots, \tilde{Y}_{N_2^{(0)}}^{(0)}; \tilde{Y}_1; \tilde{Y}_{N_1^{(1)}+1}^{(1)}, \dots, \tilde{Y}_{N_2^{(1)}}^{(1)}; \tilde{Y}_2; \dots$$

Notice that some of the sections  $\tilde{Y}_{N_1^{(r)}+1}^{(r)}, \dots, \tilde{Y}_{N_2^{(r)}}^{(r)}$  can be missing. This happens when  $\tilde{Y}_r = X_r$ .

We now set

$$\tilde{X}_{Z_r+i} := \tilde{Y}_{N_1^{(r)}+i}^{(r)}, \quad i = 1, \dots, N_2^{(r)} - N_1^{(r)}. \quad (20)$$

In view of the strong Markov property, the joint probability distribution of the random variables  $\tilde{X}$ 's defined by (17) and (20) coincides, by construction, with a finite dimensional distribution for a Markov chain with initial law  $\tilde{\pi}$  and transition matrix  $\tilde{P}$ . The random variables  $\tilde{X}$ 's have not

been defined for any time  $t \in \mathbb{N}$ . By Kolmogorov's existence theorem, we can consider however the entire chain  $\tilde{X}_0, \tilde{X}_1, \tilde{X}_2, \dots$  by suitably adding variables at the missing times. From (9) and (18), we have

$$\tilde{X}_{T_k + \sum_{r=0}^L (N_2^{(r)} - N_1^{(r)})} = k. \quad (21)$$

Furthermore, by repeating the same argument used above, we can also obtain

$$\tilde{X}_l < k, \quad (22)$$

for  $l = 0, \dots, T_k + \sum_{r=0}^L (N_2^{(r)} - N_1^{(r)}) - 1$ . Thus

$$\tilde{T}_k = T_k + \sum_{r=0}^L (N_2^{(r)} - N_1^{(r)}),$$

therefore  $\tilde{T}_k \geq T_k$ , whence the stochastic comparison in (3) follows.  $\square$

A related result is the following one, which appeared as Theorem 4.1 in [10]. Such a result gives a stronger conclusion with respect to Theorem 1 but under much stronger conditions. Its proof can be also obtained along the same line of Theorem 1.

**Theorem 2.** *Let  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$  belong to  $\Upsilon_k$  with transition matrices  $P = (p_{i,j})_{i,j \in E}$ ,  $\tilde{P} = (\tilde{p}_{i,j})_{i,j \in E}$ , and initial distributions  $\pi_0 = (\pi_0(i))_{i \in E}$ ,  $\tilde{\pi}_0 = (\tilde{\pi}_0(i))_{i \in E}$ , respectively. Under the conditions*

$$p_{i,\cdot} \succeq_{st} \tilde{p}_{i,\cdot} \text{ for each } i = 0, \dots, k-1, \quad (23)$$

$$\pi_0 \succeq_{st} \tilde{\pi}_0, \quad (24)$$

*one has the stochastic comparison*

$$T_h \preceq_{st} \tilde{T}_h, \text{ for } h = 1, \dots, k, \quad (25)$$

*where  $T_h$  is defined for  $\mathbf{X}$  in (1) and  $\tilde{T}_h$  is the analogue for  $\tilde{\mathbf{X}}$ .*

**Remark 1.** *The method of proof of Theorem 1 can be also convenient for implementation in computer programs and it is based on the skip-free property of Markov chains. A similar method of coupling had been already implemented in [7] with the aim to simulate such Markov chains and to analyze Theorem 4.1 in [10].*

*In our context, we also notice that such a method leads to efficient estimates of the difference between expected values of two different first-passage times. In fact, as can be easily proven, it is more accurate that one based on the separate estimates of the two expected values. This circumstance turns out to be useful in several situations of interest.*

*For instance, in the comparison between waiting times to words' occurrences, where the expected values can be extremely large and their differences relatively small, the estimate of expected values might reveal numerically inaccurate if compared with direct estimation of the difference between them.*

**Remark 2.** We also notice that Theorem 2 can be directly extended to the infinite case  $E = E_\infty$ . In such a frame, it can also be useful to analyze recurrence properties of Markov chains satisfying (2).

We present an example in which hypothesis of Theorem 1 is satisfied but the hypothesis of Theorem 2 fails.

**Example 1.** Let us consider  $E = \{0, 1, 2, 3\}$  and the transition matrices:

$$P = \begin{pmatrix} \frac{1}{2} + \epsilon & \frac{1}{2} + \epsilon & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \tilde{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 - \epsilon & 0 & \epsilon & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (26)$$

where  $\epsilon > 0$ . Hypothesis of Theorem 2 is never satisfied for any positive  $\epsilon$ . Taking the product of the matrices we obtain

$$P^2 = \begin{pmatrix} (\frac{1}{2} + \epsilon)^2 & \frac{1}{2} - \frac{\epsilon}{2} - \epsilon^2 & \frac{1-2\epsilon}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1+2\epsilon}{4} & \frac{1-2\epsilon}{4} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \tilde{P}^2 = \begin{pmatrix} \frac{3-2\epsilon}{4} & \frac{1}{4} & \frac{\epsilon}{2} & 0 \\ \frac{1}{2} & \frac{1-\epsilon}{2} & 0 & \frac{\epsilon}{2} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (27)$$

Therefore we can take  $m(0) = 2$  and  $m(1) = m(2) = 1$  to verify the hypothesis of Theorem 1, when  $\epsilon$  is small enough. Thus showing that  $T_k \preceq_{st} \tilde{T}_k$

The following result appears, at a first glance, to be similar to Theorem 2. However it offers a much wider range of applications. In Section 4, examples will be presented in the frame of word occurrences. Also the proof of this result can be obtained along the same line of Theorem 1, and will then be omitted.

**Theorem 3.** Given two transition matrices in the space  $\Upsilon_k$ , namely  $P = (p_{i,j} : i, j = 0, \dots, k)$  and  $\tilde{P} = (\tilde{p}_{i,j} : i, j = 0, \dots, k)$ . Let the initial measures be  $\pi = \tilde{\pi} = \delta_0$  (both the Markov chains start in zero almost surely). Suppose that there exists an integer  $m \in [1, k-1]$  such that

(i)  $T_m \succeq_{st} \tilde{T}_m$ ,

(ii) for each  $i \in [m, k-1]$   $p_{i,i+1} \leq \tilde{p}_{i,i+1}$  and  $p_{i,0} + p_{i,i+1} = 1$ .

Then  $T_i \succeq_{st} \tilde{T}_i$  for  $i \in [m, k]$ .

### 3 Asymptotic stochastic comparisons

In this section, we turn to compare the tail behaviors of two first-passage times  $T_k$  and  $\tilde{T}_k$ . The next result, in fact, aims to establishing an asymptotic form of stochastic comparison between  $T_k$  and  $\tilde{T}_k$ . From this point on in this section we can abandon the condition that the Markov chains of interest belong to the class  $\Upsilon_k$ . The following two definitions are now needed.



**Definition 1.** Given two random variables  $X$  and  $Y$  we write  $X \preceq_{a.st.} Y$  if there exists  $t_0$  such that  $P(X > t) \geq P(Y > t)$ , for each  $t \geq t_0$ .

**Definition 2.** Let  $A = (a_{i,j} : i, j = 1, \dots, k)$  and  $A' = (a'_{i,j} : i, j = 1, \dots, k)$  be stochastic matrices of a given order  $k$ . We write  $A \trianglelefteq A'$  if and only if

$$a_{i,\cdot} \preceq_{st} a'_{j,\cdot}, \quad \forall i \leq j \leq k. \quad (28)$$

**Remark 3.** The relation  $\trianglelefteq$  is stronger than (23) and it is transitive. However it is not reflexive. In this respect we have the following fact: let  $\mathbf{X}$  be an homogeneous Markov chain on the state space  $E = \{0, 1, \dots, k\}$ , with transition matrix  $P$ . The relation  $P \trianglelefteq P$  holds if and only if  $\mathbf{X}$  is stochastically monotone.

In what follows we consider that the state  $k$  is absorbing and we want to compare the asymptotic behavior of the absorbing time in the state  $k$  for two Markov chains. We also assume that the initial measure for all the chain is concentrated on the state zero.

The previous two definitions are of interest in the present context in view of the following result.

**Theorem 4.** If  $P^n \trianglelefteq \tilde{P}^n$  for each  $n$  large enough then  $T_k \preceq_{a.st.} \tilde{T}_k$ .

*Proof.* We want to check that  $P(T_k > L) \leq P(\tilde{T}_k > L)$  for  $L$  large enough. We remark, since the state  $k$  is absorbing, that the identity  $\{T_k > L\} = \{X_L \neq k\}$  holds. Then  $P_0(T_k > L) = 1 - p_{0,k}^{(L)}$ . Similarly we obtain  $P_0(\tilde{T}_k > L) = 1 - \tilde{p}_{0,k}^{(L)}$ . The thesis then follows from the inequality  $P^L \trianglelefteq \tilde{P}^L$ .  $\square$

In the following, we present some results concerning the condition that two transition matrices are such that  $P^n \trianglelefteq \tilde{P}^n$  for  $n$  large enough.

In the following result we give a probabilistic characterization of the relation  $\trianglelefteq$ .

**Lemma 1.** Let  $A, A'$  be stochastic matrices on the state space  $E = \{0, \dots, k\}$ .  $A \trianglelefteq A'$  if and only if a Markov chain  $(Z_n)_{n=0,1}$  with  $Z_n = (Y_n, Y'_n)$  on the state space  $E^2$  exists with the following properties:

- i)  $(Y_n)_{n=0,1}$  is a Markov chain with transition matrix  $A$ .  $(Y'_n)_{n=0,1}$  is a Markov chain with transition matrix  $A'$ .
- ii)  $P(Y_1 \leq Y'_1 | Y_0 = i, Y'_0 = i') = 1$ , for  $i \leq i' \in E$ .

*Proof.* Assume  $A \trianglelefteq A'$ . Set

$$Y_1(U) := \inf\{r \leq k : \sum_{l=0}^r a_{i,l} \geq U\}, \quad Y'_1(U) := \inf\{r \leq k : \sum_{l=0}^r a'_{i',l} \geq U\}, \quad (29)$$

where  $U$  is uniform distributed over  $[0, 1]$ . Then, conditionally on  $Y_0 = i$  (resp.  $Y'_0 = i'$ ),  $Y_1(U)$  (resp.  $Y'_1(U)$ ) has the law  $a_{i,\cdot}$  (resp.  $a'_{i',\cdot}$ ). Furthermore ii) holds in view of (29).

Viceversa, if i) and ii) hold, then (28) follows by definition of stochastic ordering.  $\square$

We now show that the relation  $\leq$  is maintained under products of transition matrices.

**Lemma 2.** *Let  $A, A', B, B'$  be stochastic matrices of order  $k$  such that  $A \leq A'$  and  $B \leq B'$  then  $AB \leq A'B'$ .*

An elementary but lengthy procedure can be used to prove this result. We prefer to provide a synthetic proof based on probabilistic arguments.

*Proof.* Let  $U_1$  and  $U_2$  be i.i.d. random variables uniformly distributed over  $[0, 1]$ . We consider the random variables defined by

$$Y_1(U_1) := \inf\{r \leq k : \sum_{l=0}^r a_{i,l} \geq U_1\}, \quad Y'_1(U_1) := \inf\{r \leq k : \sum_{l=0}^r a'_{i',l} \geq U_1\}, \quad (30)$$

$$Y_2(U_2) := \inf\{r \leq k : \sum_{l=0}^r b_{Y_1,l} \geq U_2\}, \quad Y'_2(U_2) := \inf\{r \leq k : \sum_{l=0}^r b'_{Y'_1,l} \geq U_2\}. \quad (31)$$

The sequence  $(Y_n)_{n=0,1,2}$  is a non-homogeneous Markov chain with transition matrix  $A$  for the first step and transition matrix  $B$  for the second step. Analogously for  $(Y'_n)_{n=0,1,2}$  with  $A'$  and  $B'$ . Now define  $X_0 = Y_0$ ,  $X_1 = Y_2$ ,  $X'_0 = Y'_0$  and  $X'_1 = Y'_2$ . These two random variables  $(X_n)_{n=0,1}$  form a Markov chain with transition matrix  $AB$ ; also  $(X'_n)_{n=0,1}$  is a Markov chain with transition matrix  $A'B'$ . The pair  $(Z_n)_{n=0,1}$  with  $Z_n = (X_n, X'_n)$  can be seen as a Markov chain on the state space  $E^2$ . In view of Lemma 1, we can conclude the proof by checking that  $P(X_1 \leq X'_1 | X_0 = i, X'_0 = i') = 1$ , for  $i \leq i' \in E$ . In fact, we have

$$\begin{aligned} P(X_1 \leq X'_1 | X_0 = i, X'_0 = i') &= P(Y_2 \leq Y'_2 | Y_0 = i, Y'_0 = i') = \\ &= \sum_{i_1 \leq i'_1} P(Y_2 \leq Y'_2, Y_1 = i_1, Y'_1 = i'_1 | Y_0 = i, Y'_0 = i'). \end{aligned}$$

Notice that, in the last equality, we are allowed to extend the sum only on  $i_1 \leq i'_1$ , in view of (30).

By the Markov property of  $Z_n$  we obtain

$$\begin{aligned} &\sum_{i_1 \leq i'_1} P(Y_2 \leq Y'_2, Y_1 = i_1, Y'_1 = i'_1 | Y_0 = i, Y'_0 = i') \\ &= \sum_{i_1 \leq i'_1} P(Y_2 \leq Y'_2 | Y_1 = i_1, Y'_1 = i'_1) P(Y_1 = i_1, Y'_1 = i'_1 | Y_0 = i, Y'_0 = i'). \end{aligned} \quad (32)$$

We also have  $P(Y_2 \leq Y'_2 | Y_1 = i_1, Y'_1 = i'_1) = 1$ , in view of Lemma 1. Therefore the r.h.s. of (32) becomes  $P(Y_1 \leq Y'_1 | Y_0 = i, Y'_0 = i')$ . The latter term is equal to 1 by (30) and this concludes the proof.  $\square$

**Remark 4.** *Lemma 2 also guarantees that  $\prod_{i=1}^n A_i \leq \prod_{i=1}^n B_i$  if  $(A_i)_{i=1,\dots,n}$  and  $(B_i)_{i=1,\dots,n}$  are stochastic matrices such that  $A_i \leq B_i$ , for  $i = 1, \dots, n$ .*

The next result has an immediate application to our problem. It in fact provides an (apparently weaker) condition sufficient for the hypothesis appearing in Theorem 4.

**Theorem 5.** Let  $P$  and  $\tilde{P}$  be stochastic matrices of order  $k$ . Assume that there exist two coprime integers  $n_1$  and  $n_2$  such that

$$P^{n_1} \leq \tilde{P}^{n_1} \text{ and } P^{n_2} \leq \tilde{P}^{n_2},$$

then

$$P^n \leq \tilde{P}^n \tag{33}$$

for  $n \geq \hat{n}(n_1, n_2) := \inf\{r : \forall r' \geq r, \quad r' = an_1 + bn_2 \text{ with } a, b \in \mathbb{N}\}$ .

*Proof.* For  $n \geq \hat{n}(n_1, n_2)$  we can write, by definition of  $\hat{n}(n_1, n_2)$ ,  $P^n = (P^{n_1})^a (P^{n_2})^b$ ,  $\tilde{P}^n = (\tilde{P}^{n_1})^a (\tilde{P}^{n_2})^b$ . Then (33) is readily obtained by Remark 4.  $\square$

We notice that the conditions given in Theorem 5 can be encountered rather often. A simple sufficient condition will be presented in the following Theorem. To this purpose we need the following notation.

Given a stochastic matrix  $P = (p_{i,j} : i, j = 0, \dots, k)$ , such that  $p_{i,k} < 1$  for  $i = 0, \dots, k-1$ , denote by  ${}_{(k)}P$  the matrix obtained from  $P$  by making  $k$  a taboo state:  ${}_{(k)}p_{i,j} = p_{i,j}/(1 - p_{i,k})$ . Let us denote by  $\lambda(P)$  the spectral gap of  $P$ .

**Theorem 6.** Let  $P$  and  $\tilde{P}$  be two stochastic matrices on the state space  $E = \{0, \dots, k\}$ . Suppose that  $p_{k,k} = \tilde{p}_{k,k} = 1$  and that  ${}_{(k)}P$  is regular. Assume furthermore that  $\lambda(P) < \lambda(\tilde{P})$ . Then there exists  $n_0$  such that  $P^n \leq \tilde{P}^n$ , for  $n \geq n_0$ .

*Proof.* First we notice that  $\lambda(\tilde{P})$  is larger than zero. Therefore, from each state  $i \in E$ , the Markov chain associated to  $\tilde{P}$  can reach the state  $k \in E$  in a finite number of steps.

In this respect we will more precisely prove that there exists  $\tilde{C} > 0$  such that the following inequality holds for any positive integer  $n$  and  $i, j \in \{0, \dots, k-1\}$

$$\tilde{p}_{i,j}^{(n)} \leq n^k \tilde{C} (1 - \lambda(\tilde{P}))^n. \tag{34}$$

Concerning the matrix  $P$  we will prove, on the other hand, that there exists  $c > 0$  such that for  $n$  large enough

$$p_{i,j}^{(n)} \geq c(1 - \lambda(P))^n. \tag{35}$$

If (34) and (35) hold, we get, for  $n$  large enough, the inequalities

$$\sum_{j=0}^l p_{i,j}^{(n)} \geq \sum_{j=0}^l \tilde{p}_{i,j}^{(n)}$$

for  $l, i, \tilde{i} \in \{0, \dots, k-1\}$ . This guarantees  $P^n \leq \tilde{P}^n$  for  $n$  large enough and concludes the proof.

In order to get the inequality in (34) we can consider the Jordan representation  $\tilde{P} = \tilde{A}^{-1} \tilde{J} \tilde{A}$  for the stochastic matrix  $\tilde{P}$ , so that we can write  $|(\tilde{J}^n)_{i,j}| \leq n^k (1 - \lambda(\tilde{P}))^n$ , for  $i, j \in \{0, \dots, k-1\}$ . By developing the products with  $\tilde{A}^{-1}$  and  $\tilde{A}$  we obtain (34) in view of the assumption  $\lambda(\tilde{P}) > 0$ .

In order to show (35) we first notice that there is only one eigenvalue of modulus  $(1 - \lambda(P))$  as a consequence of the regularity of the transition matrix  ${}_{(k)}P$ . This is an easy consequence of

Perron-Frobenius theorem, see [15]. We will denote by  $\mu$  such an eigenvalue, which is actually real (again as a consequence of Perron-Frobenius theorem). We start considering the case where  $\lambda(P) > 0$ .

In such a case, we use the Jordan representation of  $P$ . We explicitly write

$$p_{i,k}^{(n)} = \sum_{l=0}^k \sum_{m=0}^k (A^{-1})_{i,l} (J^n)_{l,m} (A)_{m,k} \quad (36)$$

To fix the ideas we consider the case in which  $(J)_{k-1,k-1} = \mu$  (the second highest eigenvalue). Furthermore we are allowed to limit attention only to indexes  $0 \leq i \leq k-1$ . In fact, for  $i = k$  the term  $p_{k,j}^{(n)}$  are zero for  $j = 0, \dots, k-1$  and one for  $j = k$ . From (36) we obtain

$$p_{i,k}^{(n)} = (A^{-1})_{i,k} A_{k,k} + (A^{-1})_{i,k-1} \mu^n (A)_{k-1,k} + o(|\mu|^n), \quad (37)$$

where  $(A^{-1})_{i,k} A_{k,k} = 1$  because  $\lambda(P) > 0$ . In this respect we claim that the products  $(A^{-1})_{i,k-1} A_{k-1,k}$  can not be all nulls. In fact, if this were the case, we would have that  $p_{i,j}^{(n)}$  does not depend on  $\mu$  which is absurd (we remember that the sum on the rows is equal to one for each  $n$ ). Therefore there exists  $\hat{i}, \hat{j} \in \{0, 1, \dots, k-1\}$  such that  $(A^{-1})_{\hat{i},k-1} A_{k-1,\hat{j}} \neq 0$  and for  $n$  large enough,

$$p_{\hat{i},\hat{j}}^{(n)} \geq \frac{1}{2} (A^{-1})_{\hat{i},k-1} \mu^n (A)_{k-1,\hat{j}} \quad (38)$$

and

$$p_{\hat{i},k}^{(n)} \leq 1 - \frac{1}{2} (A^{-1})_{\hat{i},k-1} \mu^n (A)_{k-1,k}. \quad (39)$$

Inequality (38) is guaranteed by the fact that the sum on the rows is equal to one.

Now using the regularity of the Markov chain we obtain that for  $i, j \in \{0, 1, \dots, k-1\}$

$$p_{i,j}^{(n)} \geq c (A^{-1})_{i,k-1} \mu^n (A)_{k-1,j} \quad (40)$$

where  $c$  is a positive constant. If  $\lambda(P) = 0$  means that the states  $\{0, \dots, k-1\}$  do not communicate with the state  $k$  therefore there exists an invariant measure  $\pi = (\pi_1, \pi_2, \dots, \pi_{k-1}, 0)$  with  $\pi_i > 0$  for  $i = 0, \dots, k-1$  (it is a consequence of the regularity of  $(k)P$ ). Therefore (35) is trivially satisfied. This ends the proof.  $\square$

**Remark 5.** As a consequence of Theorem 6 we obtain, for a single Markov chain with transition matrix  $P$  such that  $(k)P$  is regular, the large deviation equality  $\lim_{n \rightarrow \infty} \frac{1}{n} \ln p_{i,k}^{(n)} = \mu$  where  $\mu = 1 - \lambda(P)$ .

**Remark 6.** With obvious meaning of notation, consider the following conditions:

- a)  $\mu < \tilde{\mu}$ ;
- b)  $\exists n_0 : n > n_0, \quad P^n \leq \tilde{P}^n$ ;

$$c) T_k \preceq_{a.st.} \tilde{T}_k.$$

By summarizing Theorem 4 and Theorem 6, we have the implications  $a) \Rightarrow b)$  and  $b) \Rightarrow c)$ . It is immediate to find examples to show that both the reverse implications fail.

Concerning the interest of the relation  $\preceq$ , we also notice that it can have the following advantages with respect to the spectral-gap analysis when studying asymptotic stochastic order:

- computing products of matrices can be easier than computing eigenvalues;
- if the entries of the matrices are all rational, one can perform computations (using matrices with integer entries) without any approximation;
- one can use the same calculations to study the asymptotic and the usual stochastic orders.

## 4 Applications and occurrences of words

In this section we discuss some applications of the results of Section 2. As mentioned in the Introduction, our results can in particular be applied in the frame of words occurrences. Let  $\mathcal{A}_N \equiv \{a_1, \dots, a_N\}$  be the *alphabet* composed by the  $N$  *letters*  $a_1, \dots, a_N$ . An ordered sequence  $\mathbf{w} \equiv w_1 w_2 \dots w_k$ , where each of the elements  $w_j$  is one of the letters taken from  $\mathcal{A}_N$ , is then seen as a *word of length  $k$  on  $\mathcal{A}_N$* . We consider the space  $\mathcal{A}_N^k$  of all possible words of length  $k$  on  $\mathcal{A}_N$ .

Assume that, at any instant  $n = 1, 2, \dots$ , a letter is drawn at random from the alphabet  $\mathcal{A}_N$ . Drawings are supposed to be independent and uniformly distributed over  $\mathcal{A}_N$ . We define the space  $\Omega = \mathcal{A}_N^{\mathbb{N}}$ ; for  $\omega = (\omega_1, \omega_2, \dots) \in \Omega$ , we refer to  $\omega_n$  as *the letter at time  $n \in \mathbb{N}$* . The probability measure on  $\Omega$  is then the product measure that, at any drawing, assigns probability  $1/N$  to each letter of  $\mathcal{A}_N$ :

$$P(\omega_n = a) = \frac{1}{N}, \quad a \in \mathcal{A}_N, \quad n \in \mathbb{N}.$$

For any word  $\mathbf{w} \equiv w_1 w_2 \dots w_k$ ,  $\mathbf{w} \in \mathcal{A}_N^k$ , we consider the stopping time

$$T_{\mathbf{w}} := \inf\{n \geq k | \omega_{n-k+1} = w_1, \dots, \omega_n = w_k\},$$

i.e. the random time until the first *occurrence* of  $\mathbf{w}$ .

This scheme gives also rise to an homogeneous Markov chain  $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}}$  with state space  $E \equiv \{0, 1, \dots, k\}$ . The Markov chain  $\mathbf{X}$  is defined as follows:

- i)  $X_0 = 0$ .
- ii) For  $n \geq 1$  and  $i \in \{1, \dots, k \wedge n\}$ , one has  $X_n = i$  if
  - a.**  $\omega_{n-i+1} \omega_{n-i+2} \dots \omega_n = w_1 w_2 \dots w_i$
  - b.**  $\omega_{n-h+1} \omega_{n-h+2} \dots \omega_n \neq w_1 w_2 \dots w_h, \forall h = i+1, \dots, k \wedge n$ ;
- iii) One has  $X_n = 0$  if  $\omega_{n-i+1} \omega_{n-i+2} \dots \omega_n \neq w_1 w_2 \dots w_i$  for all  $1 \leq i \leq k \wedge n$ .

Under these positions,  $T_{\mathbf{w}}$  coincides with the time  $T_k$  of first visit to the state  $k$  for the chain. For our purposes we sometimes denote by  $P^{(\mathbf{w})} = (p_{i,j}^{(\mathbf{w})})$  the transition matrix of such a Markov chain associated to  $\mathbf{w} \in \mathcal{A}_N^k$  and denote by  $\mathcal{A}_{\mathbf{w}} = \{a \in \mathcal{A} : a \in \mathbf{w}\}$  the alphabet formed by all the distinct letters of  $\mathbf{w}$ . The alphabet  $\mathcal{A}_{\mathbf{w}}$  is then the *minimal alphabet* of  $\mathbf{w}$ .

Furthermore, for  $\mathbf{w} \equiv w_1 w_2 \dots w_k$ , we denote by  $\varepsilon_{\mathbf{w}}$  the *leading number* associated to  $\mathbf{w}$ . The latter is defined as the binary vector

$$\varepsilon_{\mathbf{w}} \equiv (\varepsilon_{\mathbf{w}}(1), \varepsilon_{\mathbf{w}}(2), \dots, \varepsilon_{\mathbf{w}}(k))$$

where each  $\varepsilon_{\mathbf{w}}(u)$  is equal to 0 or to 1, according to the following position: for  $u = 1, 2, \dots, k$

$$\varepsilon_{\mathbf{w}}(u) = \mathbf{1}\{w_{k-u+1} = w_1, \dots, w_k = \dots w_u\}.$$

Leading numbers have been introduced by J. Conway and have been repeatedly used in the applied probability literature (see in particular [12], [14] [2]), to deal with the stochastic framework described above. In particular, the distribution of  $T_{\mathbf{w}}$  only depends on the leading number  $\varepsilon_{\mathbf{w}}$  and, as a function of it, the mean value  $\mathbb{E}(T_{\mathbf{w}})$  has the explicit expression  $\mathbb{E}(T_{\mathbf{w}}) = \sum_{u=1}^k N^u \varepsilon_{\mathbf{w}}(u)$ .

In what follows, we rather analyze stochastic comparisons between the times  $T_{\mathbf{w}}$  and  $T_{\mathbf{w}'}$  of occurrences for two different words  $\mathbf{w}$  and  $\mathbf{w}'$  of the same length  $k$ . On this purpose we primarily apply the results of previous sections. We shall see furthermore that an analysis based on the leading numbers  $\varepsilon_{\mathbf{w}}$  and  $\varepsilon_{\mathbf{w}'}$  can usefully be combined with such results.

Let then  $P_{\mathcal{A}} = P_{\mathcal{A}}^{(\mathbf{w})}$  and  $P'_{\mathcal{A}} = P_{\mathcal{A}}^{(\mathbf{w}')}$  be the transition matrices corresponding to the two words  $\mathbf{w}$ ,  $\mathbf{w}'$ .

For several pairs  $\mathbf{w}$ ,  $\mathbf{w}'$ , it can happen that  $P_{\mathcal{A}}$  and  $P'_{\mathcal{A}}$  satisfy a condition of the type (23). Theorem 2 then gives us a useful criterion to check  $T_{\mathbf{w}} \preceq_{st} T_{\mathbf{w}'}$ .

The stochastic ordering  $\preceq_{st}$  is a partial order on the distributions of the times  $T_{\mathbf{w}}$ . As a first application of Theorem 2 we now show that such a partial order does admit a maximal element.

For  $a \in \mathcal{A}_N$ , let  $\underline{\mathbf{a}}$  be the word belonging to  $\mathcal{A}_N^k$  and containing all letters equal to  $a$ .

**Proposition 1.** *For any word  $\mathbf{w} \in \mathcal{A}_N^k$  and  $\underline{\mathbf{a}} \in \mathcal{A}_N^k$ , we have  $T_{\underline{\mathbf{a}}} \succeq_{st} T_{\mathbf{w}}$ .*

*Proof.* The transition probabilities for the chain associated to  $\underline{\mathbf{a}}$  are given by

$$p_{i,i+1}^{(\underline{\mathbf{a}})} = \frac{1}{N}, \quad p_{i,0}^{(\underline{\mathbf{a}})} = 1 - \frac{1}{N},$$

for  $0 \leq i \leq k-1$ . We then see that the proof is immediately obtained from Theorem 2.  $\square$

Under a simple condition, the following result shows that also a minimal element does exist w.r.t.  $\preceq_{st}$ . For  $N \geq k$ , let  $\overline{\mathbf{w}}$  be the word  $a_1 a_2 \dots a_k$ , made with the first letters of the alphabet.

**Proposition 2.** *Let  $N \geq k$ . For any word  $\mathbf{w} \in \mathcal{A}_N^k$  and for  $\overline{\mathbf{w}} \in \mathcal{A}_N^k$ , we have  $T_{\mathbf{w}} \succeq_{st} T_{\overline{\mathbf{w}}}$ .*

*Proof.* If the leading number associated to the word  $\mathbf{w} = w_1 w_2 \dots w_{k-1} w_k$  is  $(0, 0, \dots, 0, 1)$ , then there is nothing to prove because the distributions of  $T_{\mathbf{w}}$  and  $T_{\tilde{\mathbf{w}}}$  are equal. Suppose then that the leading number of the word  $\mathbf{w}$  contains more than only one 1. In such a case  $\mathbf{w}$  has a repetition of at least one letter and we can then suppose that  $a_N$ , say, is not contained among its letters, in view of the condition  $N \geq k$ . Let us consider the word  $\tilde{\mathbf{w}} = w_1 w_2 \dots w_{k-1} a_N$ . It is clear that the leading number associated with the word  $\tilde{\mathbf{w}}$  is  $(0, 0, \dots, 0, 1)$ . Therefore the distribution of  $T_{\tilde{\mathbf{w}}}$  is the same as the one of  $T_{\tilde{\mathbf{w}}}$ . Hence, in order to show the stochastic comparison  $T_{\mathbf{w}} \succeq_{st} T_{\tilde{\mathbf{w}}}$ , we prove  $T_{\mathbf{w}} \succeq_{st} T_{\tilde{\mathbf{w}}}$ . The associated Markov chains are easy to analyze because for  $i = 0, \dots, k-2$  and  $l = 0, \dots, k$  the transition probabilities verify

$$p_{i,l}^{(\mathbf{w})} = p_{i,l}^{(\tilde{\mathbf{w}})}.$$

As far as the transitions from the state  $k-1$  are concerned, we notice that for the index  $\bar{l} = \max\{l < k : \varepsilon_{\mathbf{w}}(l) = 1\} \geq 1$  we can write

$$0 = p_{k-1,\bar{l}}^{(\mathbf{w})} < p_{k-1,\bar{l}}^{(\tilde{\mathbf{w}})} = \frac{1}{N},$$

and

$$p_{k-1,l}^{(\mathbf{w})} = p_{k-1,l}^{(\tilde{\mathbf{w}})},$$

for  $l \neq 0, \bar{l}$ . Therefore

$$p_{k-1,0}^{(\mathbf{w})} = p_{k-1,0}^{(\tilde{\mathbf{w}})} + \frac{1}{N}.$$

We then see that the proof is immediately obtained from Theorem 2.

□

**Remark 7.** A same string  $\mathbf{w} \equiv w_1 w_2 \dots w_k$  can be seen as a word on different alphabets, and we must keep in mind which is the alphabet  $\mathcal{A}_N$  from which the random letters  $\omega_1, \omega_2, \dots$  are drawn. Normally, such an alphabet does not coincide with the minimal alphabet  $\mathcal{A}_{\mathbf{w}}$ . The probability distribution of  $T_{\mathbf{w}}$  depends on  $\mathbf{w}$  only through the leading number  $\varepsilon_{\mathbf{w}}$  and it depends on the alphabet  $\mathcal{A}_N$  only through its cardinality  $N$ . For brevity's sake, such dependence on  $N$  is omitted in our notation; however it cannot be neglected, generally.

In applying Theorem 2 to word occurrences, the following proposition can be of interest.

**Proposition 3.** Let condition (23) hold for  $P_{\mathcal{A}}$  and  $P'_{\mathcal{A}}$ . Then condition (23) also holds for  $P_{\hat{\mathcal{A}}}$  and  $P'_{\hat{\mathcal{A}}}$  for any alphabet  $\hat{\mathcal{A}} \supset \mathcal{A}_{\mathbf{w}} \cup \mathcal{A}_{\mathbf{w}'}$ .

*Proof.* First we notice that (23) reads

$$\sum_{l=j}^k p_{i,l} \geq \sum_{l=j}^k p'_{i,l}, \quad (41)$$

for  $i = 0, \dots, k$  and  $j = 1, \dots, k$  and that  $p_{i,l}, p'_{i,l}$  depend on the alphabet  $\mathcal{A}$  only through its cardinality  $N$ . When the alphabet  $\mathcal{A}$  is replaced by the alphabet  $\hat{\mathcal{A}}$  then each  $p_{i,l}$ , with  $l > 0$ , is replaced by  $p_{i,l}N/\hat{N}$  where  $\hat{N}$  denotes the cardinality of  $\hat{\mathcal{A}}$ . Then all the inequalities in (41) are maintained.

□

Proposition 3 guarantees the following property: once we have proved the inequality  $T_{\mathbf{w}} \preceq_{st} T_{\mathbf{w}'}$  by checking the condition (23) for a sampling alphabet, then not only  $T_{\mathbf{w}} \preceq_{st} T_{\mathbf{w}'}$  holds for any other compatible alphabet, but also it stands still on the comparison (23).

In some cases Theorem 3 can be used to compare two words  $\mathbf{w}, \tilde{\mathbf{w}}$  which cannot be compared by means of Theorem 2. An example follows.

**Example 2.** Let  $\mathbf{w}' = (A, A, B, A, A)$  and  $\mathbf{w} = (A, B, B, B, A)$  be seen as words on an alphabet with  $N \geq 5$ . By letting  $m = 4$  and by using Proposition 2 one obtains that  $T_m \preceq_{st} T'_m$ . Furthermore condition (ii) of Theorem 3 is also satisfied. Then we obtain that  $T_{\mathbf{w}} \preceq_{st} T'_{\mathbf{w}}$ . Notice that this example can be easily generalized by adding a same number  $\nu$  of letters  $A$  on the left and on the right of the two words and by adding a number  $\mu$  of letters  $B$  in the center. In other terms we are saying that, by means of Theorem 3 and Proposition 2, one can compare two words  $\mathbf{w}'$  and  $\mathbf{w}$  whose leading numbers have the form

$$\varepsilon_{\mathbf{w}}(i) = 1, \text{ for } i = 1, \dots, h; \text{ and } \varepsilon_{\mathbf{w}}(i) = 0, \text{ for } i = h + 1, \dots, k - 1,$$

$$\varepsilon_{\mathbf{w}'}(i) = 1, \text{ for } i = 1, \dots, h + 1; \text{ and } \varepsilon_{\mathbf{w}'}(i) = 0, \text{ for } i = h + 2, \dots, k - 1,$$

with  $N \geq k \geq 2(h + 1)$ .

As noticed in the Remark 3 of the previous section, the condition  $P \leq P'$  between two stochastic matrices is stronger than (23), however in order to guarantee asymptotic comparisons, we only need  $P^n \leq P'^n$ , for all  $n$  large enough. Actually, as we checked by means of several examples, the validity of the hypothesis of Theorem 5 often holds for pairs of words  $\mathbf{w}, \mathbf{w}'$  such that  $\mathbb{E}(T_{\mathbf{w}}) < \mathbb{E}(T_{\mathbf{w}'})$ .

In this respect, we can conjecture that  $\mathbb{E}(T_{\mathbf{w}}) < \mathbb{E}(T_{\mathbf{w}'})$  implies  $T_k \preceq_{a.st.} T'_k$ .

**Acknowledgments** We gratefully thank Moshe Shaked for pointing out to us relations between our results and the papers [10], [7], and [8]. We also like to thank Haijun Li for letting us see his recent paper [9].

## References

- [1] P. Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues.



- [2] R.W. Chen and A. Zame. On fair coin-tossing games. *J. Multivariate Anal.*, 9(1):150–156, 1979.
- [3] R.W. Chen, A. Zame, and B. Rosenberg. On the first occurrence of strings. *Electron. J. Combin.*, 16(1):Research Paper 29, 16, 2009.
- [4] E. De Santis and F. Spizzichino. First occurrence of a word among the elements of a finite dictionary in random sequences of letters. *Electron. J. Probab.*, 17:no. 25, 9 pp. (electronic), 2012.
- [5] A. Di Crescenzo and L.M. Ricciardi. Comparing first-passage times for semi-Markov skip-free processes. *Statist. Probab. Lett.*, 30(3): 247–256, 1996.
- [6] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.
- [7] F. Ferreira and A. Pacheco. Simulation of semi-Markov processes and Markov chains ordered in level crossing. *Next Generation Internet Networks: Traffic Engineering IEEE*, 2005:121–128 , 2005.
- [8] F. Ferreira and A. Pacheco. Level crossing ordering of skip-free-to-the-right continuous-time Markov chains. *J. Appl. Probab.*, 42(1):52–60, 2005.
- [9] H. Li. Dependence Comparison of Multivariate Extremes via Stochastic Tail Orders. Submitted
- [10] A. Irle and J. Gani. The detection of words and an ordering for Markov chains *J. Appl. Probab.*, 38A(1):66–77, 2001.
- [11] J. Keilson. *Markov chain models—rarity and exponentiality*, volume 28 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1979.
- [12] S. R. Li. A martingale approach to the study of occurrence of sequence patterns in repeated experiments. *Ann. Probab.*, 8(6):1171–1176, 1980.
- [13] A.W. Marshall and M. Shaked. New better than used processes. *Adv. in Appl. Probab.*, 15(3):601–615, 1983.
- [14] S. Robin and J. J. Daudin. Exact distribution of word occurrences in a random sequence of letters. *J. Appl. Probab.*, 36(1):179–193, 1999.
- [15] E. Seneta. *Nonnegative matrices and Markov chains*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1981.
- [16] M. Shaked and J. G. Shanthikumar. *Stochastic orders*. Springer Series in Statistics. Springer, New York, 2007.

- [17] D. Stoyan. *Comparison methods for queues and other stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, 1983.